VIRAL PEPTIDES WITH STRUCTURAL HOMOLOGY TO PROTEIN G OF RESPIRATORY SYNCYTIAL VIRUS

This invention relates to viruses of the family Paramyxoviridae, particularly viruses of the respiratory syncytial virus group. More particularly, the invention relates to the attachment protein of these viruses, and to the structure of the region of the attachment protein which is involved with binding to the cellular receptor for the virus.

BACKGROUND OF THE INVENTION

Respiratory syncytial viruses are significant pathogens of human and animals throughout the world. Virtually all humans become infected with human respiratory syncytial virus (RSV) by two years of age, and repeated infections occur throughout life. RSV is regarded as the most serious respiratory pathogen of infants and young children, but it can also cause serious disease in immunocompromised adults and in the elderly. Serious cases of infection manifest in bronchiolitis and pneumonia, and can be fatal. Estimates of the impact of RSV infection indicate that it results in 91,000 hospital admissions annually in the United States of America (Heilman, 1990) and hospitalisation of 1% of children before the age of 12 months in Britain (Cane and Pringle, 1995). Epidemics of the virus occur on an annual basis coincidental with other viruses, such as influenza and parainfluenza.

Natural immunity does not appear to provide protection against RSV infection (McIntosh and Chanock, 1990; Hall, 1994). Even infants provided with maternal antibodies are susceptible to RSV infection (McIntosh, 1990). Vaccine development strategies to combat RSV have not been successful, and in one study a formalin-inactivated virus vaccine actually exacerbated disease (McIntosh and Chanock, 1990; Hall, 1994). The only pharmaceutical agent presently available to treat RSV, Ribavirin, is expensive and complex to deliver as an

aerosol, and is of questionable efficacy (McIntosh, 1990; Levin, 1994). Thus it is apparent that a greater understanding of the infectious mechanism and immunobiology of RSV is required to develop control measures based on vaccines or antiviral agents.

RSV belongs to the *Pneumovirus* genus of the Paramyxoviridae family of single strand negative sense RNA viruses, which includes other serious pathogens such as parainfluenza, mumps and measles (McIntosh, 1990; Kingsbury, 1990) and the recently identified zoonotic, equine morbillivirus (Murray *et al*, 1995). Like other Paramyxoviridase, RSV has two membrane glycoproteins which mediate invasion of susceptible cells (Morrison and Portner, 1991). One protein, the large glycoprotein or G protein, functions in attachment to cells. The other, the so-called fusion or F protein, causes fusion between the lipid of the viral membrane envelope and the cell plasma membrane lipid bilayer. RSV infection can also be transmitted by fusion of membranes of infected cells, which have F protein expressed on their surface, with adjacent cells.

The molecular architecture of the F protein is conserved between all members of the three genera of the Paramyxoviridae; however, each genus has a characteristic attachment protein (Morrison and Portner, 1991). Members of the *Paramyxovirus* genus have attachment proteins with neuraminidase and haemagglutinating activities; the attachment proteins of the *Morbillivirus* genus are haemagglutinins, but lack neuraminidase activity; and the attachment proteins of Pneumoviruses lack both haemagglutination and neuraminidase properties. Attachment proteins of the *Paramyxovirus* (Morrison and Portner, 1991) genus participate in sialic acid receptor-type interactions, which account for their ability to agglutinate red blood cells. RSV is also reported to interact with sialic acid; however, the mechanism of RSV G protein attachment and the identity of the cellular

aerosol, and is of questionable efficacy (McIntosh, 1990; Levin, 1994). Thus it is apparent that a greater understanding of the infectious mechanism and immunobiology of RSV is required to develop control measures based on vaccines or antiviral agents.

RSV belongs to the *Pneumovirus* genus of the Paramyxoviridae family of single strand negative sense RNA viruses, which includes other serious pathogens such as parainfluenza, mumps and measles (McIntosh, 1990; Kingsbury, 1990) and the recently identified zoonotic, equine morbillivirus (Murray *et al*, 1995). Like other Paramyxoviridase, RSV has two membrane glycoproteins which mediate invasion of susceptible cells (Morrison and Portner, 1991). One protein, the large glycoprotein or G protein, functions in attachment to cells. The other, the so-called fusion or F protein, causes fusion between the lipid of the viral membrane envelope and the cell plasma membrane lipid bilayer. RSV infection can also be transmitted by fusion of membranes of infected cells, which have F protein expressed on their surface, with adjacent cells.

The molecular architecture of the F protein is conserved between all members of the three genera of the Paramyxoviridae; however, each genus has a characteristic attachment protein (Morrison and Portner, 1991). Members of the *Paramyxovirus* genus have attachment proteins with neuraminidase and haemagglutinating activities; the attachment proteins of the *Morbillivirus* genus are haemagglutinins, but lack neuraminidase activity; and the attachment proteins of Pneumoviruses lack both haemagglutination and neuraminidase properties. Attachment proteins of the *Paramyxovirus* (Morrison and Portner, 1991) generus participate in sialic acid receptor-type interactions, which account for their ability to agglutinate red blood cells. RSV is also reported to interact with sialic acid; however, the mechanism of RSV G protein attachment and the identity of the cellular

receptor for the G protein are not yet known (Markwell, 1991).

The Paramyxoviridae F protein is invariably a type I integral membrane protein, but the attachment proteins are all type II integral membrane proteins. The oligosaccharide compositions of the *Paramyxovirus* and *Morbillivirus* attachment proteins are typical of integral membrane proteins (Morrison and Portner, 1991), but the RSV attachment protein, also termed the G protein, appears to contain an unusually high proportion of carbohydrate (Morrison and Portner, 1991; reviewed by Sullender and Wertz, 1991). The gene for the RSV strain A2 attachment protein encodes a potential primary translation product of 298 amino acids with a theoretical Mr of 32588 (Satake *et al*, 1985; Wertz *et al*, 1985), but has an apparent molecular weight of 80,000-90,000 (Levine, 1977; Gruber and Levine, 1983; Lambert and Pons, 1983) as estimated by electrophoresis in polyacrylamide gels containing sodium dodecylsulfate (SDS-PAGE). The unusually high molecular weight of the RSV attachment protein as measured by SDS-PAGE has been attributed to a high content of both O- and N-linked oligosaccharides (Gruber and Levine, 1985; Lambert, 1988).

It has been suggested that a conserved region in the central part of the G protein ectodomain may be involved in ligand interactions with a cellular receptor for the G protein (Collins, 1991; Johnson *et al*, 1987); however, this remains to be demonstrated. Peptide epitope scanning experiments, using nested sets of synthetic peptides, have identified the cysteine-containing region as a subtype-specific antigenic determinant (Norrby *et al*, 1987) which shows some dependence on oxidation of the cysteines to cystine (Akerlind-Stopner *et al*, 1990). Monoclonal antibodies which react with this region also block binding of the G protein to cells (Feldman and Hendry, 1996). However, at the priority date of this

application the disulphide linkage status of this region remained to be elucidated. In addition, the glycosylation status of the serine and threonine residues in the conserved domain and around the conserved cysteine residues or the Asn-Pro-Thr sequence between cysteines 176 and 182 of the ectodomain had not been determined, and the sites of disulphide bonds were not known.

PCT/FR95/01464 by Pierre Fabre Medicament relates to a method for recombinant production of an analogue of RSV protein G, in which there is at least one modification of the amino acid sequence in a hydrophobic region of the peptide, preferably in a non-transmembrane hydrophobic region which is not essential for the biological activity of the peptide. The modifications are directed to changing the hydrophobicity of the recombinant production, either by deleting a hydrophobic amino acid of the natural sequence, or by replacing it with a non-hydrophobic amino acid. This results in the recombinant product being exposed on the membrane of the host cell by a covalent bond membrane-anchoring moiety, or in being secreted into the culture medium. One of the modifications disclosed replaces cysteine at position 173 and/or position 186 by an amino acid which cannot form a disulphide bond; this favours formation of a disulphide bond between cysteines 176 and 182, which are stated to be critical for the ~~immunogenecity~~ immunogenicity of the sequence, and avoids formation of disordered disulphide bonds. Thus while this specification does suggest that the disulphide bond between cysteines 176 and 182 is critical, it teaches away from any suggestion that a disulphide bond between cysteines 173 and 186 could be of importance. Furthermore there is no disclosure or suggestion of the presence of any glycosylation in this region. After the priority date of this application, a paper describing the 3-dimensional solution structure, as

determined by NMR spectroscopy, of a synthetic 32-residue peptide corresponding to residues 58 to 189 of the immunodominant central conserved region of protein G of bovine RSV has appeared (Doreleijers *et al*, 1996). This paper was submitted to the journal on 26 August 1996, and consequently is not prior art in respect of this application. The peptide was found to form a rigid core region comprising two short α-helices connected by type I' β-turn, and having two disulphide bridges. These were unambiguously assigned as being between cysteine 173 and cysteine 186 (outer bridge) and cysteine 176 and cysteine 182, on the basis of NMR evidence. It is stated that these disulphide bond assignments were subsequently confirmed by analysis of proteolytic digestion products and by affinity measurements (Langedijk *et al*, 1996; published after the priority date).

We have now found that the four cysteine residues of the ectodomain form a uniform disulphide bond pattern, and that the region of the protein around these residues and the conserved region lacks oligosaccharide attachment.

We have shown that peptides with these characteristics can bind to cells of a type equivalent to the *in vivo* target of RSV, and that the peptides display antiviral activity. This is to our knowledge the first time that a fragment of G protein has been shown to have any abnormal activity. This finding has significance for the design and production of vaccines, diagnostic reagents, and therapeutic compounds for human RSV and related viruses.

In particular, the absence of glycosylation leaves this region of the ectodomain open for receptor binding. Heretofore, the possibility of attached oligosaccharide being present in this region has meant that the design and interpretation of experiments to examine binding of the virus to its cellular receptor has been

exceedingly difficult, and accurate interpretation almost impossible, because of the level of variability made possible by glycosylation.

SUMMARY OF THE INVENTION

According to one aspect, the invention provides a compound having structural homology to a contiguous sequence of amino acids within the sequence representing residues 149-197 of the G protein of respiratory syncytial virus, in which

a) no oligosaccharide is linked to potential serine, threonine or asparagine attachment sites;

b) four cysteine residues are involved in disulphide linkages; and

c) the pattern of disulphide linkage is Cys 173 linked to Cys 186, and Cys 176 linked to Cys 182,

and in which said compound possesses a biological activity of respiratory syncytial virus G protein.

For the purposes of this specification, a biological activity of respiratory syncytial virus G protein is defined as one or more of

a) the ability to bind to one or more antibodies selected from the group consisting of rabbit polyclonal antibody directed against RSV, murine monoclonal antibody directed against RSV, and antibodies present in human convalescent sera from patients infected with RSV; and

b) the ability to bind to cells capable of being infected with RSV.

Preferably the virus is selected from the group consisting of human RSV subtype A, human RSV subtype B, bovine RSV, and mutants and variants thereof.

More preferably the compound is a peptide corresponding to amino acids 158 to 196 of the RSV G protein. Even more preferably the peptide corresponds to amino acids 165 to 187 of the RSV G protein.

Most preferably the compound is a peptide having one of the following amino acid sequences:

SEQ ID NO 1	KQRQNKPPSKPNNDFHFEVFNFVPCSICSNNPTCWAICKRIPNKKPGKK
SEQ ID NO 2	--------N----------
SEQ ID NO 3	----------R
SEQ ID NO 4	--H----------
SEQ ID NO 5	--------N----------
SEQ ID NO 6	--------N----------
SEQ ID NO 7	--------N----------
SEQ ID NO 8	----------R
SEQ ID NO 9	-S-SKN--K--KD-Y----------G--QL-KS---T--SN--K--
SEQ ID NO 10	-S-SKN--K--KD-Y----------G--QL-KS---T--SN--K--
SEQ ID NO 11	-P-PKN--K--KD-Y----------G--QL-KS---T--SN--K--
SEQ ID NO 12	-P-LKN--K--KD-Y----------G--QL-KS---T--SN--K--
SEQ ID NO 13	-P-LKN--K--KD-Y----------G--QL-KS---T-SSN--K--
SEQ ID NO 14	-P-LKN--K--KD-Y----------G--QL-KS---T--SN--K--
SEQ ID NO 15	-S-SKN--K--KD-Y----------G--QL-KS---T--SN--K--
SEQ ID NO 16	NPSGSI--ENHQDHNN-QTLPY----T-EG-LA-LSL-HIETERA-SRA
SEQ ID NO 17	----------P----T------R----------
SEQ ID NO 18	----------S----------R--T--------

We have additionally surprisingly found that when the cysteine residues 173 and 176 in the sequence representing residues 149-197 of the G protein of RSV are blocked, and cysteines 182 and 186 are deleted, the peptide still strongly enhances binding of ~~fluorescein~~ fluorescein-labelled peptides of sequences corresponding to residues 149-197 and 163-197 of the G protein sequence, and exerts a strong anti-viral effect. This suggests that there may be a second site for binding of G protein to cellular receptors for RSV.

Consequently, in another aspect the invention provides a compound having structural homology to a contiguous sequence of amino acids within the sequence representing residues 149-197 of the G protein of RSV, in which at least one of cysteines 173, 176, 182 and 186 is

absent or blocked, and in which said compound is not glycosylated, and has the ability to inhibit infectivity of RSV.

Any suitable assay for measuring inhibition of infectivity may be used, for example inhibition of cytopathic effect, or inhibition of viral proliferation.

Sequences encompassing conservative substitutions of amino acids are within the scope of the invention, provided that the biological activity is retained.

It is to be clearly understood that the compounds of the invention include peptide analogues, including but not limited to the following:

1. Compounds in which one or more amino acids is replaced by its corresponding D-amino acid. The skilled person will be aware that retro-inverso amino acid sequences can be synthesised by standard methods; see for example Choreo and Goodman, 1993;

2. Peptidomimetic compounds, in which the peptide bond is replaced by a structure more resistant to metabolic degradation. See for example Olson *et al*, 1993; and

3. Compounds in which individual amino acids are replaced by analogous structures for example, *gem*-diaminoalkyl groups or alkylmalonyl groups, with or without modified termini or alkyl, acyl or amine substitutions to modify their charge.

The use of such alternative structures can provide significantly longer half-life in the body, since they are more resistant to breakdown under physiological conditions.

Methods for combinatorial synthesis of peptide analogues and for screening of peptides and peptide analogues are well known in the art (see for example Gallop *et al*, 1994; Hogan, 1997). It is particularly contemplated that the compounds of the invention are useful as templates

for design and synthesis of compounds of improved activity, stability and bioavailability.

Because of the biological activity of the compounds of the invention, they are useful as therapeutic and diagnostic agents, and are also useful in screening in order to identify compounds capable of inhibiting binding of the virus to its host cell. Therefore the invention also provides

a) a diagnostic composition comprising a compound of the invention;

b) a pharmaceutical composition comprising a compound of the invention together with a pharmaceutically acceptable carrier, and optionally together with one or more other antiviral agents active against RSV;

c) a method of prevention or treatment of *Pneumovirus* infection comprising the step of administering an effective amount of a compound of the invention to a mammal in need of such treatment; and

d) a method of diagnosis of *Pneumovirus* infection, comprising exposing a biological fluid or sample from a mammal suspected of being infected with said virus to a compound of the invention, and measuring the interaction between the compound and said fluid or sample.

Diagnostic kits are also within the scope of the invention.

Because at least some compounds of the invention are immunogenic, in a further aspect the invention provides a method of immunisation against *Pneumovirus* infection, comprising the step of immunising a mammal at risk of such infection with an immunising-effective dose of a compound of the invention, said compound being immunogenic and having the ability to elicit protective antibody.

It is also contemplated that compounds of the invention may be used in conjunction with prior art vaccines. Because of the antiviral effect of compounds of the invention, this enables the dose of vaccine to be reduced, and the risk of side-effects is also reduced.

Similarly, even where the antibodies produced in response to immunisation with a compound of the invention are not protective, such antibodies will be useful as diagnostic reagents. For this application of the invention the only requirement is that the antibodies elicited have the capacity to interact with a *Pneumovirus* in a detectable manner. For example, the antibody can be coupled to a detectable marker such as a radioactive label, a fluorescent marker, a luminescent marker or an enzyme marker. The person skilled in the art will be aware of a great variety of suitable such markers. Thus both non-protective and protective antibodies directed against compounds of the invention are also within the scope of the invention.

Compounds according to the invention may be directly labelled with a detectable marker; such as those mentioned above for labelling of antibodies, and/or with a photoaffinity label, and are useful for identification and structural characterization of the cellular receptor for RSV and other Pneumonoviruses. Knowledge of the structure of the receptor and the mechanism of its interaction with the G protein is useful in the design of antiviral compounds.

The person skilled in the art will be aware that anti-idiotype antibodies directed against antibodies according to the invention provide useful structural information concerning the identity and mechanism of action of the receptor site for the G protein.

While the invention is described in detail with reference to human respiratory syncytial virus, it will be clearly understood that the invention is applicable to the genus *Pneumovirus* in general, and particularly to bovine and ovine RSVs in addition to human RSV. It will be evident from the following description that while the sequence of the G protein of bovine RSV varies to some extent, there is significant conservation of a specific sequence, and that the cysteine residues are in the same

position as in human RSV G protein. The conservation of the disulphide bond bonding pattern in all strains tested indicates that this pattern has not varied during evolution of the virus, and that it is functionally significant.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A is a diagrammatic representation of the RSV G protein predicted by gene sequence analysis. The clear area containing a single cysteine residue (I) is the cytoplasmic domain, the black region (II) is the transmembrane domain and subdomains shaded gray (III & V) are the putative heavily glycosylated regions of the ectodomain separated by the non-glycosylated disulphide subdomain (IV);

Figure 1B shows the disulphide arrangement determined in this study to involve pairing of cysteine 173 with cysteine 186, and cysteine 176 with cysteine 182;

Sub D

~~Figure 2 shows the amino acid sequence~~ encompassing residues 149-197 of the G proteins of variants of different subtypes of RSV. Sequences 1-15 are human RSV strains, sequence 1 is that of the A2 strain of the A subtype (Satake *et al*, 1985; Wertz *et al*, 1985), sequence 2 is the Long A strain of the A subtype (Johnson *et al*, 1987), and sequences 3-8 are natural variants of the A subtype isolated in the same locality in a single year (Cane *et al*, 1991). Sequences 9-15 are natural variants of the B subtype isolated in different localities over a 29-year period (Johnson *et al*, 1987; Cane *et al*, 1991; Sullender *et al*, 1990; Sullender *et al*, 1991). Sequence 16 is that of Bovine RSV (Lerch *et al*, 1990). Sequences 17 and 18 are variants of human RSV, R10c/1 and R10c/10, which were generated by propagation of the Long A strain in the presence of a monoclonal antibody directed to the cysteine-containing constant region of the ectodomain of the G protein (Rueda *et al*, 1994);

Figure 3 illustrates the HPLC separation of protease fragments of RSV strain A2 G protein produced by

tryptic digestion of the entire protein (A) and by digestion of the fraction eluting at approximately 73 minutes during HPLC of the tryptic digest with different proteases (B-D). Chromatograms B, C and D are of digests obtained using pepsin, thermolysin, and post-proline cleavage enzyme, respectively;

Figure 4 shows MALDI-TOF-MS spectra of components of the fraction eluting at approximately 73 minutes during HPLC of the tryptic digest of Figure 3A. A spectrum of an aliquot of the tryptic fraction is shown in A, and a spectrum of an aliquot of the unfractionated peptic digest of this tryptic fraction is shown in B. The unfractionated peptic digestion was performed as for Figure 3, except that the temperature was 22°C. C and D represent spectra of fractions eluting at approximately 61 minutes and 64 minutes during HPLC of the peptic digest (Figure 3B). Spectrum B was recorded in the linear mode. Spectra A and B were recorded with matrix 3 and spectra C and D were recorded with matrix 4;

Sub. D2

~~Figure 5 shows the proposed identities of peptide fragments detected by MALDI-TOF-MS in various digests and HPLC fractions. Theoretical *m/z* values corresponding to the proposed fragment identities are presented next to the corresponding sequence. All *m/z* values are for the oxidized sequences, except for fragments 1R, 2R and 3R, which are for reduced forms of these sequences;~~

Figure 6 illustrates MALDI-TOF-MS spectra obtained by reduction of fragments 3(A) and 2(B) of the peptic digest of tryptic fragment 1. Matrix 4 was used to record both spectra;

Figure 7 shows MALDI-TOF-MS spectra of an unfractionated thermolytic digest of tryptic fragment 1(A) and the thermolytic fragment eluting at approximately 54 minutes during HPLC of the thermolytic digest of Figure 3C (B). Conditions for thermolytic digestion were as described for Figure 3 except that 0.5 mM $CaCl_2$ was used. Matrix 4 was used to record both spectra;

Figure 8 shows MALDI-TOF-MS spectra of an unfractionated post-proline cleavage enzyme digest of tryptic digest 1 (A) , the fraction eluting at approximately 63 minutes during HPLC (Figure 3D) of this digest (B) and an unfractionated post-proline cleavage enzyme digest of peptic fragment 2 (C). The unfractionated digests were prepared as for Figure 3, except that final enzyme concentrations of 0.1 mg/ml and 10 µg/ml were used to obtain spectra A and C, respectively. Matrix 4 was used to record all spectra;

Figure 9 shows post-source decay fragment ion spectra of peptic fragment 2 (A), peptic fragment 3 (B) and post-proline cleavage enzyme fragment 5 (C). Fragment ions 1-7, inclusive, correspond to an N-terminal fragment ion series of the b-type resulting in fragmentation of the amino acid residues with or without the peptide, A I C K. Fragment ions 8-14, inclusive, correspond to a C-terminal fragment ion series of the y-type. This fragmentation pattern is presented in a diagrammatic form in Figure 10 together with *m/z* values for the numbered fragment ions of peptic fragment 2. An expanded portion of spectrum A (*m/z* = approximately 850 to 1650) is presented above the full spectrum. Matrix 3 was used to record spectra A and B and matrix 1 was used to record spectrum C; and

Sub. D3

~~Figure 10 illustrates the proposed fragmentation pattern of peptic fragment 2 based on data from Figure 9A.~~

Sub. D4

~~Figure 11 shows the sequences of residues 149-197~~ from human, ~~bovine,~~ and ovine RSV G protein, indicating the ~~features which are common to all strains.~~

Sub. D5

~~Figure 12 shows the sequences of the peptide derivatives described herein.~~

The cysteinyl residues of all but peptide 4 were oxidised to cystine residues with the same linkage arrangement as in the native G protein.

The cysteinyl residues of peptide 4 were retained in a form protected with the acetamidomethyl functionality. Ac denotes an acetylated amino terminus.

Amide denotes a carboxyl terminal amide.

fitc denotes presence of fluorescein isothiocarbamyl β-alanine at the amino terminus.

bb denotes derivatisation of the amino terminus as a benzoyl benzylamide.

biot denotes derivatisation of the amino terminus as a biotinyl amide.

Figure 13 shows flow cytometry analysis of binding of fluoresceinyl-149-197 to HEp-2 cells. Relative fluorescence of cells is plotted on the abscissa versus relative cell number on the ordinate after incubation of cells (1) without or with (2) 468nM, (3) 1.17μM (4) 2.34μM, (5) 4.68μM, (6) 11.72μM, (7) 23.44μM and (8) 46.88μM fluorescent peptide.

Figure 14 shows flow cytometry analysis of binding of fluoresceinyl-163-197 to HEp-2 cells and the effect of non-fluorescent peptide derivatives on cell bound fluorescence. Relative fluorescence of cells is plotted on the abscissa versus relative cell number on the ordinate after incubation of cells with fluorescent peptide at 890nM (1) in the absence of non-fluorescent peptide derivatives and in the presence of (2) 4.89μM Ac149-197, (3) 19.6μM Ac163-197 and (4) 22.46μM Ac149-177.

Figure 15 shows flow cytometry analysis of binding of fluoresceinyl-149-197 to HEp-2 cells and the effect of Ac149-197 on cell bound fluorescence. Relative fluorescence of cells is plotted on the abscissa versus relative cell number on the ordinate after incubation of cells (1) without fluorescent peptide and with fluorescent peptide at 468nM (2) in the absence of non-fluorescent peptide derivatives and (3) in the presence of 4.89μM Ac149-197.

Figure 16 shows flow cytometry analysis of binding of fluoresceinyl-149-197 to HEp-2 cells and the effect of Ac149-197 oxidised A and B chains of insulin on cell bound fluorescence. Relative fluorescence of cells is plotted on the abscissa versus relative cell number on the

ordinate after incubation of cells (1) without fluorescent peptide and with fluorescent peptide at 455nM (2) in the presence of 31.6μM oxidised B chain of insulin, (3) in the presence of 22.2μM oxidised A chain of insulin, (4) in the absence of non-fluorescent peptide derivatives and (5) in the presence of 4.89μM Ac149-197.

Figure 17 shows flow cytometry analysis of binding of fluoresceinyl-VTRQRRARNGASRS to HEp-2 cells and the effect of Ac149-197 on cell bound fluorescence. Relative fluorescence of cells is plotted on the abscissa versus relative cell number on the ordinate after incubation of cells (1) without fluorescent peptide and with fluorescent peptide at 8μM (2) in the absence of non-fluorescent peptide derivatives and (3) in the presence of 4.89μM Ac149-197.

Figure 18 shows flow cytometry analysis of binding of fluoresceinyl-149-197 to HEp-2 cells and the effect of non-fluorescent truncated peptide derivatives of the 149-197 sequence on cell bound fluorescence. Relative fluorescence of cells is plotted on the abscissa versus relative cell number on the ordinate after incubation of cells (1) without fluorescent peptide and with fluorescent peptide at 455nM (2) in the absence of non-fluorescent peptide derivatives and in the presence of (3) 12.2μM Ac149-197, (4) 48.87μM Ac163-197, (5) 41.26μM Ac149-190 and (6) 56.1μM Ac149-177.

Figure 19 shows flow cytometry analysis of binding of fluoresceinyl-149-197 to HEp-2 cells and the effect of benzoylbenzyl-149-197 on cell bound fluorescence. Relative fluorescence of cells is plotted on the abscissa versus relative cell number on the ordinate after incubation of cells (1) without fluorescent peptide and with fluorescent peptide at 468nM (2) in the absence of non-fluorescent peptide derivatives and (3) in the presence of 4.1μM benzoylbenzyl-149-197.

Figure 20 shows confocal scanning microscopy of HEp-2 cells incubated with 4.68μM fluoresceinyl-149-197 without other peptide derivatives.

Figure 21 shows confocal scanning microscopy of HEp-2 cells incubated with 4.68μM fluoresceinyl-149-197 in the presence of 3.1μM Ac149-197.

Figure 22 shows confocal scanning microscopy of HEp-2 cells incubated with 4.68μM fluoresceinyl-149-197 in the presence of 22.4μM Ac149-177.

DETAILED DESCRIPTION OF THE INVENTION

Inspection of the deduced amino acid sequence of the G protein of many RSV isolates (Satake *et al*, 1985; Wertz *et al*, 1985; Johnson *et al*, 1987; Cane *et al*, 1991; Sullender *et al*, 1990; Sullender *et al*, 1991; Garcia *et al*, 1994) reveals several structural domains, which are illustrated in Figure 1. The N-terminal region (residues 1-38) is located on the inner aspect of the viral envelope, and is relatively conserved, as is the transmembrane region (residues 39-66). However, the ectodomain (232 residues) has two regions of comparative variation bordering a central region (residues 149-197), which is conserved within subgroups and contains 4 closely positioned cysteine residues which are conserved in all RSV sequences. As shown in Figure 2, this region also has a sequence of 13 amino acids, including 2 of the conserved cysteine residues, which is identical in all wild type isolates of RSV that infect humans. The variable regions contain potential sites for N-linked glycosylation of asparagine, and have an abundance of serine and threonine residues which are potential sites for O-linked oligosaccharides. Thus the ectodomain comprises 7 occurrences of the consensus tripeptide sequence asparagine-Xaa-threonine/serine, the motif for asparagine-linked glycosylation, although at three of these sites Xaa is proline, which is a contraindication of such glycosylation. A relative abundance of proline residues in

the regions high in serine and threonine suggests the presence of O-linked oligosaccharides. Some functional studies indicate an active role for the O- and N-linked oligosaccharides of the RSV membrane proteins in cellular invasion (Lambert, 1988), but systematic analyses of the structures and positioning of the oligosaccharides remain to be performed.

Experimental Procedures

Peptide Isolation

Strain A2 RSV G protein was isolated by immunoaffinity chromatography (Walsh, 1984), with modifications to include immunoaffinity columns specific for RSV F and nucleocapsid proteins prior to a final G protein antibody column and elution from the affinity column with potassium thiocyanate. Lyophilized G protein samples were reconstituted in sufficient 0.1 M NH_4HCO_3 to result in a final concentration of 0.01% Triton X-100 (v/v), 0.14 M NaCl and 10 mM phosphate buffer (pH 7.2 in the absence of NH_4HCO_3). N-ethylmaleimide was also added to a final concentration of 1 mM. Digestion of the intact protein was performed for 4 hours at 37°C using two additions of 1% (w/w) of sequencing grade trypsin (Boehringer-Mannheim) with the second addition made at 2 hours. Subdigestion of peptide fragments isolated by high performance liquid chromatography (HPLC) was achieved with pepsin (Boehringer-Mannheim), thermolysin (Calbiochem) and/or post-proline cleavage enzyme (Seikagaku Corporation, Tokyo). Details of individual subdigestion protocols are described below in association with specific experiments.

Proteolytic fragments were isolated by reverse phase HPLC (RP-HPLC) using slight variations of a previously described protocol (Gorman *et al*, 1990), using a 2.1 mm x 25 cm column of octadecasilica (Vydac), a flow rate of 150 µl/min and a linear gradient from 0.1% (v/v) aqueous trifluoroacetic acid to 80% (v/v) aqueous CH_3CN containing 0.09% (v/v) trifluoroacetic acid, developed over

90 minutes. Gradients were generated using a Hewlett Packard chromatography system comprising a 1090M solvent delivery system under the control of a DOS Chemstation, and elution of peptides was monitored at 214 nm using a 1090 diode array detector.

Peptide Derivatives

Peptide derivatives 1-8 (Figure 12) were synthesised and purified on a contract basis by Auspep Pty. Ltd., Parkville, Australia. The peptidyl moiety of derivative 9 was synthesised and purified in our laboratory. Derivatives 1 and 3 were synthesised by both t-Boc/Benzyl and FMoc/t-Butyl solid phase strategies. Derivative 2 was synthesised exclusively by the t-Boc/Benzyl based solid phase strategy, and other peptides were produced by the FMoc/t-Butyl solid phase strategy only.

Amino-terminal derivatisation of peptides 1-8 was conducted while the peptides were resin bound, with all side chain protecting groups intact. Acetylation was achieved by acylation with acetic anhydride under basic conditions. Benzoylbenzoic acid and biotin were coupled using standard coupling reactions. Fluorescent derivatives 5 and 6 were produced by coupling FMoc-β-alanine to the amino-terminal residue of the nominated sequence and subsequently reacting the deprotected amino terminus of the β-alanine residues with fluorescein-isothiocyanate under basic conditions. Peptide derivatives were cleaved from their resin supports under standard acidolytic cleavage conditions prior to subjecting the cysteinyl residues of derivatives 1-3 and 5-8, inclusive, to air oxidation in NH_4HCO_3. Oxidation was monitored by electrospray ionisation mass spectrometry, and allowed to proceed until the mass of the crude products diminished by 4 Daltons, which indicated the formation of 2 disulphide bonds from the 4 cysteinyl residues of these derivatives. Derivatives 4 and 9 were not subjected to air oxidation, since the cysteinyl residues of

5 were maintained in side- chain protected form as acetamidomethyl derivatives, and 9 lacked cysteinyl residues. Derivative 9 was subjected to amino-terminal labelling by direct reaction with fluorescein-isothiocyanate in an aqueous medium under alkaline conditions.

All peptide derivatives were purified by reverse phase high performance liquid chromatography to greater than 80% purity, as indicated by integration of peaks in the chromatogram. In all cases the desired peptide derivative was shown to be the predominant product by electrospray ionisation mass spectrometry.

Peptide derivatives were dissolved in deionised dististilled water for use in flow cytometry and confocal scanning microscopy, or dissolved in tissue culture medium and sterile filtered for antiviral assays.

Mass Spectrometry

Matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF-MS) was performed using a Bruker Reflex mass spectrometer (Bruker-Franzen Analytik GmbH, Bremen, Germany). Samples were deposited on to target surfaces after mixing with an equal volume of the supernatant fraction of a saturated mixture of α-cyano-4-hydroxy-cinnamic acid (matrix 1) or 3,5-dimethoxy-4-hydroxy-cinnamic acid (matrix 2) in 33% (v/v) aqueous CH_3CN containing 0.1% (v/v) trifluoroacetic acid (Beavis *et al*, 1992), or a 10 mg/ml solution of α-cyano-4-hydroxy-cinnamic acid in 50% (v/v) C_2H_5OH/CH_3CN (matrix 3), or a 10 mg/ml solution of 2,6-dihydroxyacetophenone in 50% (v/v) C_2H_5OH/CH_3CN containing 0.1 M di-ammonium hydrogen citrate (matrix 4) (Gorman *et al*, 1996). Ionization was achieved using a nitrogen laser pulsed at a repetition rate of 3 Hz. Laser irradiance was adjusted to threshold levels in order to observe intact molecular ions, which were accelerated to a potential of 28.5 kV on to a hybrid microchannel plate-photomultiplier linear detector or subsequently reflected

with a reflectron potential of 30 kV on to a dual microchannel plate detector. Except where specifically noted, molecular ion masses were determined in the reflectron mode. Matrix ions were deflected by application of a 2 kV potential across deflector plates placed immediately after the ion acceleration region in order to avoid saturation of the detectors. All measurements were made at a digitization rate of 250 Mhz. Masses were assigned to intact peptide ions by reference to an external calibration file created using the flight times of the components of a mixture of 200 fmoles of angiotensin II (MH^+ = 1046.56; monoisotopic mass), 400 fmoles of adrenocorticotropic hormone residues 18-39 (MH^+ = 2466.73; average mass) and 2 pmoles of bovine insulin (MH^+ = 5734.54; average mass) applied to a separate target spot.

Analysis of metastable ions arising from post-source decay was performed using 25% stepwise decrements in the reflectron potential and increasing the laser irradiance to optimise production of ions in each voltage window (Spengler *et al*, 1992; Kaufmann *et al*, 1993; Kaufmann *et al*, 1994). Masses were assigned to metastable ions by reference to a calibration table created by determining the behaviour of metastable ions of known mass, produced from adrenocorticotropic hormone residues 18-39 , at various reflectron potentials (Rouse *et al*, 1995). As described above, data were acquired at a digitization rate of 250 MHz. Assembly of the individual spectra for each reflection voltage on to a continuous mass scale was performed using Bruker FAST software routines within the Bruker XMASS software package.

Reduction of Disulphide Bonds

Disulphide containing peptides were reduced, after adjusting HPLC fractions to pH 5 by the addition of 1 M aqueous di-ammonium hydrogen citrate to a final concentration of 0.1 M, by addition of 50 mM aqueous

tris(2-carboxyethyl)-phosphine (Molecular Probes) to a final concentration of 5 mM and incubating the mixtures at 65°C for 20 minutes. Reduced samples were mixed with an equal volume of 2,6-dihydroxyacetophenone in 50% (v/v) C_2H_5OH/CH_3CN and applied to a sample target for mass spectrometric analysis.

Edman Degradation

Stepwise amino acid sequence analysis of peptides was performed by automated Edman degradation using a Hewlett Packard G1000A solid-phase protein sequenator.

Cells and Viruses

Human Laryngeal tumor cells (HEp-2) were grown as monolayers in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) foetal bovine serum. Cells were prepared for flow cytometry, confocal scanning microscopy and electron microscopy by washing twice with phosphate-buffered saline (PBS) and detached by incubation with 1mM EDTA in PBS for 20 minutes at 37°C. The cells were suspended by agitation and centrifuged at 1000 x g for 5 minutes, then washed with PBS by centrifugation and resuspended in PBS to approximately $1x10^6$ cells/ml. Between 2 and 22µl of solutions of non-fluorescent and/or fluorescent peptide derivatives were incubated at 0°C for 1 hr with 2.5 X 10^5 detached cells in 250µl for these assays. These incubation conditions were not detrimental to the viability or morphology of the cells. After incubation, cells were pelleted by low speed centrifugation, the supernatants were decanted and the cells resuspended in 250µl PBS at 0°C.

Flow Cytometry

Propidium iodide was added to the cell suspension to a concentration of 8µg/ml immediately before the cells were analysed by flow cytometry using a Coulter EPICS® Elite flow cytometer. The illuminating wavelength was

488nm, and forward light scatter, 90° light scatter (side scatter), and fluorescence emission at 525nm (fluorescein) and at >600nm (propidium iodide) were recorded. The fluorescence of peptide derivatives bound to the live single cell population was recorded by gating on forward and side light scatter, and gating on cells negative for propidium iodide fluorescence.

Confocal Scanning Microscopy

Confocal scanning microscopy was carried out on unfixed cells remaining following flow cytometric analysis. The cells were mounted in PBS without the addition of anti-fade compounds, and observed using a Bio-Rad MRC500 confocal microscope with illumination from an argon-ion laser and a 40x0.7 NA objective . The standard fluorescein (green) fluorescence filter set was employed. Images were accumulated using Kalman filtering.

Example 1 Disulphide Determination by Analysis of Proteolytic Fragments

The whole G protein was digested with trypsin, and a fraction of the tryptic digest was then subjected to further digestion with pepsin, thermolysin, or post-proline cleavage enzyme as described. The digests were analysed by RP-HPLC, and the results are summarized in Figure 3.

Chromatogram A resulted from injection of 100 µl of a digest which contained approximately 40 µg of G protein exposed to trypsin for 4 hours at 37°C (see experimental procedures for detailed conditions). Subdigestions of the fraction eluting at 73 minutes in chromatogram A were achieved after removal of the CH_3CN using a stream of high purity nitrogen. Pepsin (10 µl of a 1 mg/ml solution in 5% (v/v) formic acid) was added to 100 µl of the fraction and digestion was allowed to proceed for 2 h at 37°C prior to injecting 100 µl of the digest to generate chromatogram B. The fraction was prepared for thermolytic digestion by mixing 45 µl of the fraction with

45 µl of 0.1 M NH_4HCO_3 and 11 µl of 0.01 M $CaCl_2$. Thremolysin (10 µl of a 1 mg/ml solution in 0.1 M NH_4HCO_3) was added and digestion allowed to proceed for 2 hours at 37°C prior to injecting 100 µl to generate chromatogram C. The fraction was prepared for post-proline cleavage enzyme digestion by mixing 45 µl of the fraction with 20 µl of 0.1 M NH_4HCO_3 before adding 45 µl of 0.1 mg/ml solution of the enzyme in the enzyme in 0.1 M ammonium acetate to give a final pH of 6.5. Digestion was allowed to proceed for 2 hours at 37°C prior to injecting 100 µl of the digest to generate chromatogram D.

The following fragments were produced by enzymatic digestion:

Fragment 1	Trypsin cleavage of G protein;
Fragments 2 & 3	Pepsin cleavage of Fragment 1;
Fragment 4	Thermolysin cleavage of Fragment 1;
Fragment 5	Cleavage of Fragment 1 with post-proline protease;
Fragment 6	Cleavage of Fragment 2 with post-proline protease;
Fragment 2R	Reduced Fragment 2; and
Fragment 3R	Reduced Fragment 3.

Cleavage with Trypsin

RP-HPLC of tryptic digests of the G protein consistently revealed only a few discrete peaks of absorbance at 214 nm at the column breakthrough and at approximately 10.5, 25 and 73 minutes (Figure 3A). The remainder of the chromatogram consisted of several series of broad baseline rises consisting of poorly resolved peaks of absorbance at 214 nm. This appearance is consistent with extensive attachment of heterogenous oligosaccharides to most of the peptide constituents of the digest. The peak eluting at 10.5 minutes represented N-ethylmaleimide, which was added to the digest as a precaution against disulphide bond interchange or disulphide bond oxidation of

cysteine-containing fragments. The peak eluting at 10.5 minutes was not present in a digest from which N-ethylmaleimide was omitted.

The other early-eluting discrete peaks failed to produce data when examined by MALDI-TOF-MS; however, as shown in Figure 4A, the peak at 73 minutes produced intense ion signals at *m/z* values of 4108.0 and 4125.3. These masses are consistent with the tryptic peptide spanning residues 152 to 187 of the G protein (Fragment 1), taking into account oxidation of the four cysteines of this sequence to cystine residues, which leads to a loss of 4 Da in mass from both peptides. Partial conversion of the N-terminal glutamine residue 152 to pyroglutamic acid apparently accounts for the difference of 17 Da in mass between these ions. These data also indicate that Fragment 1 does not have any carbohydrate attached to the asparagine residue at position 179, the threonine residue at position 181, or the serine residues at positions 157, 174 and 177.

Subcleavage with Pepsin

Cleavage of isolated Fragment 1 with pepsin and analysis of the unfractionated digest of MALDI-TOF-MS produced ions consistent with residues 152 to 164 and residues 169 to 187 (Figure 4B). Ions at *m/z* = 1653.2 and *m/z* = 1670.0, representing residues 152 to 165, were spaced by 17 Da, indicating a mixture of glutaminyl and pyroglutamyl residues derived from the N-terminus of the original tryptic Fragment 1. Ions at 2096.7 and 2115.7 represented the C-terminal portion of Fragment 1 (residues 169 to 187), allowing for peptide bond cleavage within a cystine loop of the heavier peptide (Fragments 2 and 3). As for the parent tryptic peptide, 4 Da had to be subtracted from the theoretical masses of the C-terminal portions to obtain correspondence with the experimentally observed masses. This provided further evidence that the

four cysteine residues of the sequence were in disulphide linkage.

Fragments 2 and 3 were isolated by RP-HPLC (Figure 3B) and analysed as isolated fragments which provide corroboration of the identities assigned to ions in the unfractionated digest, especially the spacing of 18 Da of ions representing the C-terminus of tryptic Fragment 1, which indicated intra-disulphide loop peptide bond cleavage for the heavier ion. Peaks at 52 and 52.5 minutes corresponded to N-terminal peptic fragments, the peak of 61 minutes corresponded to the heavier of the C-terminal peptic fragments (Fragment 2; Figure 4C), and the peak at 64 minutes corresponded to the lighter of the C-terminal fragments (Fragment 3; Figure 4D). Mass analysis of these peptides after chemical reduction showed an increase in mass of 4 Da for the lighter of the C-terminal fragments, consistent with acquisition of four protons as a consequence of reduction of two disulphide bonds (Figure 6A). The heavier of the unreduced peptides actually lost mass, corresponding to loss of the sequence Ala-Ile-Cys-Lys and gain of 3 protons following reduction of two disulphide bonds (Figure 6B). These data indicate that the peptic cleavage within a disulphide loop to produce the heavier of the C-terminal peptic fragments (Fragment 2) occurred after the sole tryptophan of the parent tryptic Fragment 1 (tryptophan 183 of the G protein).

Subcleavage with Thermolysin

Cleavage of the tryptic Fragment 1 with thermolysin and direct analysis of the digest reaction by MALDI-TOF-MS produced ions consistent with cleavages prior to phenylalanine residue 19 of Fragment 1 and prior to isoleucines at positions 24 and 34 of Fragment 1. Ions diagnostic of these cleavages were at *m/z* values of 912.9 and 1106.9 (Figure 7A). The ion at *m/z* = 912.9 is consistent with disulphide bridging between 173 and 186

(residues 19-23 of Fragment 1 linked to residues 34-36), and the ion at *m/z* = 1106.9 is consistent with disulphide bridging between cysteines 176 and 182, but without cleavage of intra-cystine-loop peptide bonds. RP-HPLC of the thermolytic digest revealed a complex series of peaks, as shown in Figure 3C. Most of these peaks appeared to have been derived from the enzyme preparation, as they were present in a chromatogram produced using a mock digest which lacked Fragment 1. Some of these peaks produced ions which did not correspond to any portion of Fragment 1. The digest had one peak at approximately 38.5 minutes which yielded an ion at *m/z* = 912.2, consistent with the cystene 173 to 186 loop. However, this fraction was a mixture of peptides, as evidenced by other ions at higher *m/z* values, and was not characterized any further. The fraction eluting at approximately 54 minutes was not present in the mock digest, and yielded the correct mass for the peptide containing the cysteine 176 to 182 loop (Fragment 4), as shown in Figure 7B. Automated Edman degradation sequencing produced an N-terminal sequence of Ile-Xaa-Ser-Asn, which is consistent with the identification of this fragment by mass analysis, with Xaa representing a gap corresponding to retention of the PTH derivative of one half cystine residue on the sequencer cartridge due to disulphide linkage to its half cystine pair.

Subcleavage with Post-Proline Cleavage Enzyme

Cleavage of tryptic Fragment 1 with a post-proline cleavage enzyme produced a fragment (Fragment 5) containing the two disulphides, but without intra-disulphide loop cleavage at the proline residue in the cystine 176 to 182 loop. Ions with *m/z* values indicative of this fragment (*m/z* = 1639.7) were observed in both the digestion reaction mixture (Figure 8A) and in a fraction at approximately 63 minutes generated by HPLC of the reaction mixture (Figure 3D). This is shown in Figure 8B. Post-

proline cleavage of the heavier of the C-terminal peptic fragments of Fragment 1 (*ie.* Fragment 2) also achieved cleavage at the extra-cystine loop proline but not the intra-loop proline (Fragment 6), as indicated by appearance of an ion at m/z = 1658.8 in the digest illustrated in Figure 8C. We attempted to obtain further fragments consistent with the disulphide bond pattern indicated above, using cleavage strategies involving chymotrypsin, proteinase K and mild acid; however, these strategies either failed to contribute additional data or were not rewarding at all.

The residues involved in each fragment, and their observed and theoretical m/z ratios, are summarized in Table 1, and the structure of each fragment, indicating its disulphide bonds, is shown in Figure 5. It can be seen that the pattern of disulphide bonding is the same in each fragment, and that this pattern is destroyed by reduction.

Table 1

Enzymatically-Produced Fragments of the RSV G-Protein Containing Ectodomain Cystine/Cysteine Residues

Fragment	Cleavage	Residues	Observed/*m/z*	Theoretical *m/z*
1	Trypsin cleavage of G protein	152 - 187	4125.3/4108.0	4124.7/4107.7[1,2]
2	Pepsin cleavage of 1	169 - 187	2115.8	2115.5[1,3]
3	Pepsin cleavage of 1	169 - 187	2097.9	2097.5[1]
4	Thermolysin cleavage of 1	175 - 184	1106.9	1107.3[1]
5	Post-Proline Protease cleavage of 1	173 - 187	1639.7	1639.9[1]
6	Post-Proline Protease cleavage of 2	173 - 187	1658.8	1658.0[1,3]
2R	Reduction of 2	169 - 183	1685.7	1685.9[3]
3R	Reduction of 3	169 - 187	2101.3	2101.5

[1] Calculated assuming that all four cysteine residues are involved in disulphide bonds

[2] The difference of 17 Da between these ions is accounted for by cyclization of the N-terminal glutamine residue to pyroglutamic acid.

[3] Calculated based on the addition of 18 Da to unreduced fragments to account for cleavage at the C-terminal peptide bond of tryptophan 183 and loss of the disulphide-linked A I C K sequence upon reduction.

Example 2 Disulphide Determination by Mass Spectrometric Based Sequence Analysis of Fragments 2, 3 and 5

Analysis of products of gas-phase metastable decomposition of ions of peptic Fragment 2 by MALDI-TOF-MS produced a series of fragment ions illustrated in Figure 9A, which were consistent with the disulphide bonding pattern deduced by analysis of proteolytic fragments (Figure 5). The series of fragment ions obtained is represented both diagrammatically and in tabular form in Figure 10. These fragment ions are of the b- and y-type (Roepstorff and Fohlman, 1984) representing cleavages at the peptide bonds along the peptide backbone. Fragment ions 1 to 4 inclusive are a series of b-type ions which are independent of possible disulphide bonding arrangements, but fragment ions 5 to 7 inclusive are diagnostic of the indicated disulphide linkage, due to the concomitant loss of mass of the A I C K sequence together with fragmentations at Cys173 or Ser174 or Ile175 of the larger peptide chain. Furthermore, this diagnosis is supported by the occurrence of fragment ions 8 to 11, which also bear the mass of the A I C K sequence, and by the failure to observe this mass accompanying fragment ions 12 to 14.

A comparable analysis of peptic Fragment 3 revealed the sequence specific b-type fragment ions 2, 3 and 4 seen for peptic Fragment 2, which are independent of the disulphide bonding pattern which are shown in Figure 9B. Fragment ions of Fragment 3 at *m/z* values of 1983.3, 1836.7 and 1735.6 are potentially equivalent to y-type ions 8, 9 and 10 respectively of Fragment 2 when a mass difference due to inclusion of an additional peptide bond in Fragment 3 is taken into account. Other prominent fragment ions of Fragment 3 were apparent at *m/z* values of 288.4, 648.9 and 719.1. Only the ions at *m/z* values of 648.9 and 719.1 were also seen with Fragment 2. The ion at *m/z* = 648.9 can be rationalised as a b-type ion resulting

from cleavage between Ser174 and Ile175 of the peptide, plus cleavage of the disulphide bond involving Cys173. The ions at *m/z* values of 288.4 and 719.1 cannot be easily accounted for. None of the fragment ions of Fragment 2 which were used to support the disulphide bonding arrangement were observed with Fragment 3. This supports the logic used in interpretation of the ion series of Fragment 2 used to define the disulphide pattern. This logic was dependent upon identifying ions as being combinations of metastable ion masses due to cleavage along the peptide backbone at the N-terminus of the Fragment 2 plus mass contributed by the disulphide-linked A I C K sequence. The A I C K sequence was also linked by a peptide bond to Trp183 in Fragment 3, which explains why it was not liberated together with metastable fragment ions from this peptic fragment.

As shown in Figure 9C, post-proline cleavage Fragment 5 failed to produce metastable fragment ions of comparable intensity to those produced by Fragments 2 and 3. These observations with Fragment 5 support the conclusions drawn from the data on Fragments 2 and 3. Fragment 5 does not have the N-terminal N F V P sequence, or the cleavage following Trp183 required to produce fragment ions equivalent to fragment ions 1 to 4 and 8 to 11 seen with Fragments 2 and 3.

As shown in Figure 10 and Table 2, the results of analysis of metastable ions produced by peptic Fragments 2 and 3 and post-proline cleavage Fragment 5 (Figure 9) were consistent with the pattern deduced by analysis of proteolytic fragments as shown in Figure 5.

Values in Table 2 for the observed fragment ions are an average of three separate determinations with peptic Fragment 2. Fragment ions 5-7 (b-type ions), inclusive, and 8-11 (y-type ions), inclusive, are diagnostic of the proposed disulphide pattern, because they account for the

mass of the linked peptide, AICK, in addition to the mass produced by the cleavage of amino acids as indicated.

Table 2

N-Terminal Fragments			C-Terminal Fragments		
	Observed	Predicted		Observed	Predicted
1	-	115.10	8	2001.2 ± 0.8	2001.44
2	262.4 ± 0.2	262.27	9	1854.3 ± 1.1	1854.26
3	361.4 ± 0.3	361.41	10	1755.2 ± 1	1755.13
4	458.6 ± 0.2	458.52	11	1658.9 ± 1	1648.01
5	992.9 ± 0.5	993.25	12	1123.1 ± 0.3	1123.29
6	1080.6 ± 0.2	1080.33	13	1036.4 ± 0.6	1036.21
7	1193.5 ± 0.6	1193.49	14	923.1 ± 0.6	923.05

Example 3 Synthesis of Disulphide-Bonded Peptide

A fully synthetic peptide of sequence corresponding to that of Peptide 1 in Figure 2 (SEQ ID NO: 1), *ie.* amino acids 149 to 197 inclusive of the G protein of RSV strain A2, was prepared using conventional solid-phase methods.

This peptide preferentially formed the same disulphide bonding pattern as that identified in the previous examples.

Example 4 Binding of RSV G Protein Residues 149-197 to RSV Susceptible Cells

A series of peptide derivatives was produced based on the amino acid sequence and disulphide bond configuration of residues 149-197 of human RSV A2 strain G protein. These peptide derivatives also have features common to all strains of human RSV, bovine RSV and ovine RSV. These are shown in Figure 12.

Binding experiments were conducted using RSV-susceptible HEp-2 cells and fluoresceinyl peptide derivatives of the 149-197 sequence, in the presence and absence of non-fluorescent peptide derivatives with acetyl or benzoylybenzyl substituents on their amino termini. Binding of fluoresceinyl peptide derivatives was assessed using flow cytometry and confocal scanning microscopy.

Peptides corresponding to residues 149-197 and 163-197 of the A2 strain of human RSV, extended at their amino termini by adding β-alanine followed by reaction with fluorescein isothiocyanate, were assesed for their capacity to bind to uninfected HEp-2 cells using flow cytometry. Non-adherent HEp-2 cells exhibited a pronounced increase in fluorescence after incubation with fluoreceinyl-149-197 and fluoresceinyl-163-197 followed by removal of the supernatant by centrifugation. The degree of fluorescence increase was dependent upon fluoresceinyl-149-197 peptide concentration over the range of 468nM to 46.8μM, but saturation was not observed in this concentration range, as

shown in Figure 13. Cells incubated with at 890nM and 8.9μM fluoresceinyl-163-197 showed a comparable degree of fluorescence increase as with the fluorescent derivative of 149-197(Figure 14).

Non-fluorescent residues Ac149-197 (peptide derivative 1) failed to prevent binding of fluoresceinyl-149-197 (Figure 15) and fluoresceinyl-163-197 (Figure 14). On the contrary, the binding of the fluorescent peptide derivatives, over the concentration range of 468nM to 8.9μM, appeared to be enhanced in the presence of 4.9μM Ac149-197 (Figures 14 and 15). Enhancement of fluorescence of bound fluoresceinyl-149-197 was not caused by control peptides such as the isolated oxidised A (22.2μM) or B (31.6μM) chains of insulin (Figure 16). Unrelated fluorescein-labelled peptide derivative 9 produced a modest increase in fluorescence of HEp-2 cells (Figure 17) at 8μM; however, the intensity of fluorescence was at least an order of magnitude less than with comparable concentrations of fluoresceinyl-149-197 or fluoresceinyl-163-197. The intensity of bound unrelated fluorescent peptide was not substantially increased by the inclusion of 4.89μM Ac149-197 (Figure 17).

Peptide derivatives synthesised with truncations to the N- and C-termini of the 149-197 sequence and derivatised by acetylation of their N-termini (Figure 12) also enhanced fluorescence of fluoresceinyl-163-197 (Figure 14) and fluoresceinyl-149-197 (Figure 18) bound to HEp-2 cells when added together with the fluorescent derivative. While there were variations in absolute fluorescence values of experiments performed on different days, the relative enhancing effects of the peptide derivatives was consistently in the order 4>2>3>1. In addition to causing the most pronounced enhancement of fluorescence, the acetamidomethyl peptide derivative 5 (Figure 12) also caused aggregation of cells (Figures 14 and 18), and membrane permeability not observed with other peptide derivatives.Derivative 7, with benzoylbenzoic acid

on the N-terminus of residues 149-197, was also found to enhance the fluorescence of HEp-2 cell bound fluoresceinyl-149-197 (Figure 19).

Example 5 Binding and Capping of Peptides

Binding of fluoresceinyl-149-197 to HEp-2 cells was also evident by confocal scanning microscopy in the form of peripheral fluorescence on the plasma membrane of the cells. In the absence of non-fluorescent peptide derivatives this fluorescence was in the form of patches(Figure 20). However, the fluorescence was distributed into larger assemblies when Ac149-197 was added to the cells with fluoresceinyl-149-197 under conditions that enhance HEp-2 cell fluorescence in flow cytometry assays. These larger assemblies appeared to form caps with a polar distribution on the surface of the cells, and a degree of cell aggregation was evident (Figure 21). Confocal microscopy was not conducted with addition of all fluorescent derivatives, but the dramatic assembling effect was also evident with Ac149-177 (Figure 22), which caused the most pronounced fluorescence enhancement detectable by flow cytometry. The aggregation of cells by Ac149-177 seen by flow cytometry was also evident by confocal microscopy (Figure 22).

The enhancement of fluorescence of HEp-2 cell bound fluoresceinyl-149-197 and fluoresceinyl-163-197 by non-fluorescent derivatives containing various portions of the 149-197 sequence is evidence of a complex binding interaction between this region of the RSV G protein and site(s) on cellular receptor(s). This may indicate an interaction involving cooperative formation of productive ligand interaction(s) or low affinity binding that is shifted toward the bound state by the increase in ligand concentration caused by inclusion of the non-fluorescent peptide derivatives. The transition of the fluorescence distribution of HEp-2 cell bound derivatives of 149-197 from a patch appearance to a more dense cap appearance in

the presence of non-fluorescent derivatives containing various portions of the 149-197 sequence is further evidence of specific ligand-receptor sites on the 149-197 region of the RSV G protein.

The demonstration that Ac149-177 exhibited a profound enhancement of HEp-2 cell bound fluorescence, capping of cell bound fluorescence distribution shows that a binding interaction site is located in the region of the RSV G protein between residues 149-177. This surprising finding also indicates that the influence of the binding site of Ac149-177 does not depend upon disulphide bonds, since the two cysteinyl thiols of this peptide derivative were protected by the acetamidomethyl group. However, a binding interaction may also occur between cellular receptor(s) and the 149-197 region of the RSV G protein via another complementary site that depends upon the disulphide bond configuration defined above for the G protein ~~in~~

Benzoylbenzyl-149-197 enhanced the fluorescence of HEp-2 cell bound fluoresceinyl-149-197 in a similar manner to Ac149-197, indicating that the benzoylbenzyl derivative interacted with cellular receptor(s) and that this derivative is useful for photocrosslinking studies aimed at identifying the cellular receptors for the RSV G protein.

Example 6 Antiviral Activities of Synthetic Peptide Derivatives

The ability of peptide derivatives to inhibit the cytopathic effect (cpe) of RSV on HEp-2 cells was used to assess whether the peptide derivatives bind to cellular receptors for RSV in a biologically relevant manner.

The impact of peptide derivatives on the cytopathic effect (cpe) of the A2 strain of human RSV on HEp-2 cells was assessed using monolayers grown in 96-well plastic tissue culture plates. Serial two-fold dilutions of peptide derivatives were added to cell monolayers in 50µl of sterile tissue culture medium prior to incubation

for 1-1.5 hr at 0°C. Virus (50 µl) was then added and incubation continued for 1 hr at 0°C, followed by 4 days at 37°C without removal of excess peptide derivatives or virus. Monolayers were then fixed with formalin and stained with neutral red. Inhibition of cpe was determined by comparison with control cells infected with virus in the absence of any peptide derivatives. Monolayers were then fixed with formalin and viable cells stained with neutral red.

The binding of fluoresceinyl-149-197 and fluoresceinyl-163-197 to HEp-2 cells, demonstrated by flow cytometry and by confocal microscopy in the form of patches on the plasma membrane, shows that specific ligand binding interaction site(s) for cellular receptor(s) are contained within this region of the RSV G protein.

Peptide derivatives 1-4 (Figure 12) inhibited the cytopathic effect (cpe) of the A2 strain of human RSV on HEp-2 cells to different extents. The IC50 values for Ac149-177 and Ac149-190 were approximately 5-10 µM, which were comparatively more effective than the other peptide derivatives, which had IC50 values of approximately 50 µM. Oxidised A and B chains of insulin failed to inhibit the cpe of RSV on HEp-2 cells at 28 an 40 µM, respectively. Oxidised A and B chains of insulin failed to inhibit RSV-induced cpe of HEp-2 cells when included in these assays at 28 and 40 µM, respectively.

DISCUSSION

We have now shown that the disulphide bonding pattern of the ectodomain of the unusual attachment protein or G protein of RSV involves a preferred stable configuration with Cys173 linked to Cys186, and Cys176 linked to Cys182. This was achieved by a combination of analysis of proteolytic fragments of the protein and further analysis of metastable ions produced from the proteolytic fragments during MALDI-TOF-MS. These findings represent a potent demonstration of the utility of

MALDI-TOF-MS for the mass analysis and structural elucidation of peptides, with simultaneous characterization of post-translational modifications. The observation that the synthetic peptide corresponding to residues 149 to 197 formed the same disulphide bond arrangement as the viral protein strongly indicates that the preferred configuration is actually the sole configuration

It is apparent that the disulphide loop between Cys176 and Cys182 forms a restricted region of accessibility, since the proline residue in this loop was not vulnerable to post-proline cleavage enzyme under conditions where another proline residue preceding the disulphide loops was cleaved. In contrast the loop between Cys173 and Cys186 appears to be more accessible, since peptide bonds within this loop were cleaved by both pepsin and thermolysin.

Our results are surprising in view of the extremely high degree of post-translational glycosylation of the RSV G protein. The characteristics of the chromatogram obtained by HPLC of a tryptic digest of the G protein indicated extensive glycosylation of the ectodomain. However, it is evident that the region of the G protein ectodomain which we have defined does not carry any oligosaccharides. Furthermore, the 3 and 10 residues attached to the N- and C-terminal ends, respectively, of this region do not have the potential for glycosylation. This represents 49 of the 232 residues of the ectodomain, or approximately 20%, which are not glycosylated.

Thus it is apparent that the ectodomain of the G protein has a subdomain structure, in which two highly glycosylated subdomains of 83 and 101 amino acid residues are separated by a comparatively smaller non-glycosylated subdomain which has a highly defined disulphide bond arrangement. The occupancy status of the remaining potential glycosylation sites and the characteristics of the glycans are yet to be determined. Definition of the glycosylation of the subdomains is essential in order to

assess the contribution of oligosaccharides to the mechanism of action and immunobiology of the G protein

The region of the G protein ectodomain containing the disulphides and the peptide sequences immediately adjacent to the N- and C-terminal ends of this region appear to have functional significance for both receptor interactions and immunological reactivity of the G protein. A reassessment of earlier antigenic analyses in the light of our results indicates an important role for the disulphides in maintaining the structural integrity of the protein. Studies with nested sets of synthetic peptides representing overlapping portions of the ectodomain have demonstrated that rabbit polyclonal antibodies and murine monoclonal antibodies to the G protein and human convalescent sera from natural infection all react in common with a peptide containing three of the four cysteines of the ectodomain (Norrby *et al*, 1987). The rabbit antisera also reacted with a variety of peptides, but the convalescent sera only reacted with two other peptides, one of which overlapped the commonly reactive peptide and another closely positioned peptide, while the monoclonals only reacted with the commonly reactive cysteine containing peptide. It is possible that a wider spectrum of antibody reactivities with the G protein might have been evident had the epitope scanning experiments utilised glycosylated domains of the G protein.

Subsequent studies indicated that the cysteine containing region formed a subgroup-specific antigenic determinant, and that intact disulphide bonds were important for this characteristic (Akerlind-Stopner *et al*, 1990). Further support for the immunological importance of this region came from studies with escape mutants generated using a neutralising monoclonal antibody (Rueda *et al*, 1994). These escape mutants had mutations at either Cys182 or Cys186 of their G proteins (Figure 2). The overall effect of these changes was apparently sufficient to enable the mutant viruses to escape neutralisation but to retain

functionally effective G proteins. Thus, it is possible that the mutant viruses had subtle, but immunochemically significant, differences in the surface chemistry of the cysteine regions of the ectodomains of their G proteins, while retaining a functionally competent structural fold. From our results it is apparent that these mutants retained the ability to form one of the two correct disulphides of the ectodomain, that is either the Cys173 to Cys186 or the Cys176 to Cys182 linkage. The replacement of the cysteine residues by arginine residues in both cases presumably compensated for the loss of stability associated with loss of a disulphide bond by replacement with a residue wit the ability to form a salt bridge.

The functional importance of the disulphide region is shown by studies using various polypeptides produced by recombinant routes or by chemical synthesis to immunize experimental animals against challenge with live RSV. A recombinant vaccinia virus expressing a polypeptide encompassing the disulphide region (residues 1-230 of the G protein) has been shown to produce neutralising antibodies and to confer protection from challenge with live RSV, with a response equivalent to that elicited by a recombinant vaccinia virus expressing full length G protein. In contrast, a recombinant vaccinia virus expressing a polypeptide terminating at residue 180 (residues 1-180) failed to provide protection (Olmstead *et al*, 1989). Another recombinant vaccinia construct encompassing residues 124-203, also conferred protection (Simard *et al*, 1995). The importance the disulphide region is also illustrated by finding that a synthetic peptide containing three of the cysteine residues (residues 174-187), which conferred protection from challenge by live RSV despite the fact that only non-neutralising serum antibodies were produced (Trudel *et al*, 1991).

Nested sets of synthetic peptides have also been used to attempt to define the portion of the G protein

which interacts with the cellular receptor for RSV (Feldman and Hendry, 1996); however, none of these peptides from the ectodomain blocked binding of the G protein to RSV-susceptible cells. It was postulated that the peptides may have lacked secondary or tertiary structural elements required for interaction with receptors. However, as with the immunological studies, the lack of oligosaccharides on these peptides appears to have been overlooked by previous workers. Furthermore, the disulphide bond status of these peptides does not appear to have been addressed. It is conceivable that the structural elements missing from the synthetic peptides involve correct disulphide bond pairing or another element contained by such disulphide bonds.

Knowledge of the actual disulphide linkage pattern of G protein ectodomain has facilitated experimental assessment of the functional and immunological roles of the disulphides and neighbouring peptide sequences. Our results have enabled us to develop strategies for synthesis of peptides with the correct disulphide bridging to probe receptor binding and immunological interactions of this portion of the G protein.

The antiviral activity of the non-fluorescent peptide derivatives containing various regions of the 149-197 sequence, including Ac149-177 with acetamidomethyl protected cysteinyl residues, shows that the binding of fluoresceinyl and biotinyl peptide derivatives evident by flow cytometry and electron microscopy, and the influences of non-fluorescent peptide derivatives on this binding, represent biologically relevant interaction(s) between the RSV G protein and cellular receptor(s) for the virus. Furthermore, these antiviral activities indicate that mimics of the structures of these peptide derivatives, including non-peptide compounds, and/or the actual peptide derivatives described herein and/or peptide derivatives containing various regions of the peptide derivatives

described herein may form the basis of therapeutic control of RSV infections.

The demonstration that derivatives of the sequence including residues 149-197 of the G protein of RSV have the appropriate structure to bind to RSV susceptible cells in a manner that reflects the biologically relevant event of binding of virions to cells shows that these derivatives and/or mimics of the structures of these derivatives are suitable components of vaccines for immunotherapeutic control of RSV.

Furthermore, we have demonstrated that synthetic strategies for this subdomain of the G protein did not need to consider glycosylation. Peptide by-products with the incorrect disulphide bonding arrangements are theoretically possible from synthetic approaches; however, these can easily be identified by routine methods, and will serve as control peptides for assessment of the relevance of the particular disulphide pattern determined herein. In fact, no such peptide by-products were observed. Peptide products with residual blocking groups were obtained, and can also be used as controls.

It will be apparent to the person skilled in the art that while the invention has been described in some detail for the purposes of clarity and understanding, various modifications and alterations to the embodiments and methods described herein may be made without departing from the scope of the inventive concept disclosed in this specification.

References cited herein are listed on the following pages, and are incorporated herein by this reference.

REFERENCES

Åkerlind-Stopner, B., Utter, G., Mufson, M.A., Orvell, C., Lerner, R.A. and Norrby, E.
Journal of Virology, 1990 64 5143-5148

Alansari H. and Potgieter, N.D.
Virology, 1993 196 873-877

Beavis, R.C., Chaudhary, T. and Chait, B.T.
Organic Mass Spectrometry, 1992 27 156-158

Cane, P.A., Matthews, D.A. and Pringle, C.R.
Journal of General Virology, 1991 72 2091-2096

Cane P.A. and Pringle, C.R.
Seminars in Virology, 1995 6 371-378

Choreo and Goodman
Acc. Chem. Res., 1993 26 266-273

Collins, P.L.
The Paramyxoviruses. Kingsbury, D.W. ed New York: Plenum Press; 1991 103-161

Doreleijers, J.F., Langedijk, P.M., Hård, K., Boelens, R., Rullmann, J.A.C., Schaaper, W.M., van Oirschot, J.T. and Kaptein, R.
Biochemistry, 1996 35 14684-14688

Feldman S.A. and Hendry, R.M.
Proceedings of the 13th Annual Meeting of the American Society for Virology, 1996

Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.P.A. and Gordon, E.M.
J. Med. Chem., 1994 37 1233-1251

Garcia, O., Martin, M., Dopazo, J., Arbiza, J., Frabasile, S., Russi, J., Hortal, M., Perez-Brena, P., Martinez, I., Garcia-Barreno, B. and Melero, J.A.
Journal of Virology, 1994 68 5448-5459

Gorman, J.J., Corino, G.L. and Shiell, B.J.
Biomed. Environ. Mass Spectrometry, 1992 27 156-158

Gorman, J.J., Ferguson, B.L. and Nguyen, T.N.
Rapid Communications in Mass Spectrometry, 1996 10 (5) 529-536

Gruber C. and Levine, S.
Journal of General Virology, 1983 64 825-832

Gruber C. and Levine, S.
Journal of General Virology, 1985 66 417-432

Hall, C.B.
Science, 1994 265 1393-1394

Heilman, C.A.
Journal of Infectious Diseases, 1990 161 402-406

Henderson, F.W., Collier, A.M., Clyde Jr, W.A. and Denny, F.W.
New England Journal of Medicine, 1979 300 530

Hogan, Jr., J.C.
Nature Biotechnology, 1997 15 328-330

Johnson, P.R., Spriggs, M.K., Olmsted, R.A. and Collins, P.L.
Proc. Natl. Acad. Sci. USA, 187 84 5625-5629

Kaufmann, R., Kirsch, D. and Spengler, B.
International Journal of Mass Spectrometry and Ion Processes, 1994 131 355-385

Kaufmann, R., Spengler, B. and Lutzenkirchen, F.
Rapid Communications in Mass Spectrometry, 1993 7 902-910

Kingsbury, D.W.
Virology. Fields, B.N./ Knipe, D.M. ed New York: Raven Press, 1990 945-962

Lambert, D.M.
Virology, 1988 164 458-466

Lambert D.M. and Pons, M.W.
Virology, 1983 130 204-214

Langedijk, J.P.M., Schaaper, w.M.M., Meloen, R.H. and van Oirschot, J.T.
J. Gen. Virol., 1996 77 1249-1257

Lerch, R.A., Anderson, K. and Wertz, G.W.
Journal of Virology, 1990 64 5559-5569

Levin, M.J.
Journal of Paediatrics, 1994 124 S22-S27

Levine, S.
Journal of Virology, 1977 21 427-431

McIntosh, K and Chanock, R.M.
in Virology: Fields, B.N. and Knipe, D.M. (eds) Raven Press NY (1990)

Markwell, M.A.
The Paramyxoviruses. Kingsbury, D.W. ed New York: Plenum Press; 1991 407-426

Morrison T. and Portner, A.
The Paramyxoviruses. Kingsbury, D.W. ed New York: Plenum Press; 1991 347-382

Murray, J., Selleck, P., Hooper, P., Hyatt, A., Gould, A., Gleeson, L., Westbury, H., Hiley, L., Selvey, L., Rodwell, B., and Ketterer, P.
Science, 1995 268 94-97

Norrby, E., Mufson, M.A., Alexander, H., Houghten, R.A. and Lerner, R.A.
Proc. Natl. Acad. Sci. USA, 987 84 6572-6576

Olmstead, R.A., Murphy, B.R., Lawrence, L.A., Elango, N., Moss, B. and Collins, P.L.
Journal of Virology, 1989 63 (1) 411-420

Olson *et al*
J. Med. Chem., 1993 36 3039-3049

Roepstorff, P. and Fohlman, J.
Biomedical Mass Spectrometry, 1984 11 601

Rouse, J.C., Yu., W. and Martin S.A.
Journal of American Society for Mass Spectrometry, 1995 6 822-835

Rueda, P., Garcia-Barreno, B. and Molero, J.A.
Virology, 1994 198 653-662

Satake, M., Coligan, J.E., Elango, N., Norrby, E. and Venkatesan, S.
Nucleic Acids Research, 1985 21 7795-7812

Simard, C., Nadon, F., Seguin, C. and Trudel, M.
Antiviral Research, 1995 28 303-315

Spengler, B., Kirsch, D., Kaufmann, R. and Jaeger, E.
Rapid Communications in Mass Spectrometry, 1992 6 105-108

Sullender, W.M., Anderson, K. and Wertz, G.W.
Virology, 1990 178 195-203

Sullender, W.M. Mufson, M.A., Anderson, L.A. and
Wertz, G.W.
Journal of Virology, 1991 65 5425-5434

Sullender, W.M. and Wertz, G.W.
The Paramyxoviruses. Kingsbury, D.W. ed. New York: Plenum
Press, 1991 383-406

Trudel, M., Nadon, F., Seguin, C. and Binz, H.
Virology, 1991 185 749-757

Walsh
Journal of General Virology, 1984 65 761-767

Wertz, G.W., Collins, P.L., Huang, Y., Gruber, C.,
Levine, S. and Ball, L.A.
Proc. Natl. Acad. Sci. USA, 1985 82 4075-4079

Sub. C13